Robust Distribution-Based Winsorization in Composite Indicators Construction

Boudt, Kris; Todorov, Valentin; Wang, Wenjing

published in
Social Indicators Research
2020

DOI (link to publisher)
10.1007/s11205-019-02259-w

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Boudt, K., Todorov, V., & Wang, W. (2020). Robust Distribution-Based Winsorization in Composite Indicators Construction. *Social Indicators Research*, *149*(2), 375-397. https://doi.org/10.1007/s11205-019-02259-w

Social Indicators Research (2020) 149:375–397
https://doi.org/10.1007/s11205-019-02259-w

ORIGINAL RESEARCH

Robust Distribution-Based Winsorization in Composite Indicators Construction

Kris Boudt[1,2,3] · **Valentin Todorov**[4] · **Wenjing Wang**[2,5]

Accepted: 27 December 2019 / Published online: 4 January 2020

Abstract

Composite indicators are widely used to determine the ranking of countries, organizations or individuals in terms of overall performance on multiple criteria. Their calculation requires standardization of the individual statistical criteria and aggregation of the standardized indicators. These operations introduce a potential propagation effect of extreme values on the calculation of the composite indicator of all entities. In this paper, we propose robust composite indicators for which this propagation effect is limited. The approach uses winsorization based on a robust estimate of the distribution of the sub-indicators. It is designed such that the winsorization affects only the composite indicator rank but has no effect on the entities ranking in each sub-indicator. The simulation study documents the benefits of distribution-based winsorization in the presence of outliers. It leads to a ranking that is closer to the clean data ranking when compared to the ranking obtained using either no winsorization or the traditional winsorization based on empirical quantiles. In the empirical application, we illustrate the use of winsorization for ranking countries based on the United Nations Industrial Development Organization's Competitive Industrial Performance index. We show that even though the sub-indicator ranking does not change, the robust winsorization approach has a material impact on the ranking of the composite indicator for countries with large discrepancies in the scores of the sub-indicators.

The views expressed herein are those of the author(s) and do not necessarily reflect the views of the United Nations Industrial Development Organization.

✉ Wenjing Wang
wenjingwang@cufe.edu.cn

Kris Boudt
kris.boudt@ugent.be

Valentin Todorov
v.todorov@unido.org

1 Department of Economics, Ghent University, 9000 Ghent, Belgium

2 Solvay Business School, Vrije Universiteit Brussel, 1050 Brussel, Belgium

3 School of Business and Economics, Vrije Universiteit Amsterdam, 1081 Amsterdam, the Netherlands

4 United Nations Industrial Development Organization (UNIDO), 1400 Vienna, Austria

5 School of Statistics and Mathematics, Central University of Finance and Economics, 100081 Beijing, China

Keywords Composite indicator · Invariance · Robust estimation · Winsorization

1 Introduction

According to the Organisation for Economic Co-operation and Development (OECD) *Glossary of Statistical Terms*, "a composite indicator is formed when individual indicators are compiled into a single index on the basis of an underlying model of the multi-dimensional concept that is being measured" (OECD 2008a). Composite indicators are widely used to evaluate and rank the performance of entities by producing an aggregate ordinal or cardinal measure of multi-dimensional data. Governments and international organizations are using composite indicators for analysis and advocacy due to its simplicity for being understood (Maggino 2017; Saltelli 2007). Their composite nature necessitates standardizing the individual indicators prior to aggregation. Popular choices include the use of so-called z-score or min-max normalization. The former is popular in combination with arithmetic aggregation, while the latter is often used in conjunction with geometric aggregation (OECD 2008b).

The z-score and the min-max normalization approaches have in common that, because of the standardization, the presence of an extreme observation leads to an explosion of the scale statistics (standard deviation, range) and thus to an implosion to zero of most observations in the standardized series of that indicator. From the viewpoint of ranking using individual indicators, this phenomenon naturally has no effect on the ranking. As we will show, however, it can have large effects on the ranking obtained using the composite indicator.

In this paper, we classify a composite indicator as robust when the occurrence of an extreme value in the data has only minor effects on the ranking.[1] We document in detail the robustness issue in standard approaches to ranking using composite indicators, and we propose alternatives that employ winsorization to compute composite indicators that are less sensitive to extreme observations in the data.[2]

Winsorization of data consists of shrinking the extreme observations to a boundary value. Its implementation requires statistical rules in order to detect the extreme observations and to calibrate the boundary value. The most common winsorization approach sets the boundary value to a low (or high) percentile of the empirical distribution and detects outliers as those observations exceeding the boundary value (see, e.g., Saisana 2010).

[1] A related but different literature on composite indicator robustness studies the sensitivity of the composite indicator-based ranking to the choice of standardization, aggregation and imputation methods. Booysen (2002) provides an overview of dimensions to classify and evaluate the various implementations. Saisana and Saltelli (2011) use Monte Carlo techniques to quantify the sensitivity of the composite indicator rankings to the choice of standardization, aggregation and imputation method. Their technique is used in UNIDO (2013) to evaluate the robustness of the Composite Industrial Performance index. Davino and Romano (2014) describe how multivariate statistical techniques such as analysis of variance (ANOVA) and principal components analysis (PCA) can be used to analyze the variability due to the different composite indicator construction methods. Cherchye et al. (2008) propose a linear programming technique to obtain a robust ranking, which holds for a wide set of normalization and/or aggregation procedures. Permanyer (2011) studies the lack of ranking robustness by choosing specific weights scheme for the variables included in composite indicators.

[2] Alternative approaches to reduce the sensitivity of composite indicators to outliers include the use of data transformation (e.g., taking logarithms, Box-Cox function) (see, e.g., Correia 2014; OECD 2008b; Parente 2019) and data trimming (see, e.g., Cherchye et al. 2007; Saisana and Saltelli 2008).

This approach is not robust when the observation corresponding to the empirical quantile used is also an outlier. A further problem is that all observations exceeding the boundary value receive the same value, implying that the winsorized data are no longer informative about the original ranking of the indicator data. We propose a solution in which a robustly estimated distribution is used when winsorizing the data. The novelty in the proposed approach is to set the boundary value using a percentile of a robust fit of the distribution of indicators and to let the replacement values vary across observations such that the ranking of the winsorized indicator equals the ranking at the original indicator.

We document the good properties using an extensive simulation study that compares the proposed approach to the standard empirical quantile-based winsorization method in the presence of outliers. The comparison is based on the average rank shift indicator that compares the estimated ranking with the ranking obtained using outlier-free data. We find that the proposed robust multivariate distribution-based winsorization approach demonstrates the lowest average rank shift in the presence of extreme values as compared to the other methods.

In the empirical application, we illustrate the results of the proposed winsorization approach by applying it to the Competitive Industrial Performance (CIP) and the Green Industrial Performance (GIP) composite indicators by UNIDO (Moll de Alba and Todorov 2018; UNIDO 2002, 2013, 2017). We find that winsorization has a material effect in the case of countries with large discrepancies in the performance of the indicators. For the CIP, the most notable case is China, which performs best in two indicators but ranks only 41st in another indicator. Winsorizing the best performing indicators using our approach sets them at values by which China still has the largest value for the corresponding individual indicator, but its composite indicator drops in rank from 3rd to 17th.

For the GIP, we find that an extreme value in the CO_2 emission indicator data for Syria leads to an explosion of the max-min statistic and hence an implosion of the normalized indicator data. The winsorization leads to composite indicators that can resist this implosion effect. The normalized CO_2 emission indicator is no longer shrunk to zero and thus has a material effect on the composite indicator. As a consequence, the composite GIP indicator ranks for Ukraine and Viet Nam drop from 55 and 61 to 69 and 73, respectively.

The remainder of the paper is organized as follows. Section 2 describes our motivation. Section 3 presents the proposed methodology. Section 4 presents the results of the analysis based on simulated data, while Sect. 5 presents the application result of the CIP and GIP data. Section 6 contains our conclusions.

2 Motivating Example

2.1 Definitions

Composite indicators are useful for their ability to integrate large amount of information into easily understood formats and are valued as a suitable tool for ranking all entities in a population according to their performance in various domains. However, the ranking robustness of a composite indicator may suffer from the presence of the extreme values in one or more indicators.

To compare the rank robustness, we use average rank shifts (ARS) to measure the rank difference between two versions of a composite indicator for n entities. The average rank shift is defined as

Table 1 Numeric example to illustrate the impact of extreme values on the ranking of composite indicators constructed by z-score or min-max normalization and linear aggregation. Results with winsorization are also shown to demonstrate the effectiveness of the method. ARS stands for average rank shift with respect to the rank based on the clean data

ID	Clean data				Contaminated outlier				Winsorized contaminated data			
	I^1	I^2	CI	Rank	I^1	I^2	CI	Rank	$w_{DB}(I^1)$	$w_{DB}(I^2)$	CI	Rank
Panel A: z-score normalization and linear aggregation												
1	1.68	1.05	0.44	1	1.68	1.05	− 0.49	5	1.68	1.05	− 0.09	3
2	1.02	1.87	0.34	2	1.02	1.87	0.71	2	1.02	1.50	0.20	2
3	1.45	1.07	− 0.01	3	1.45	1.07	− 0.50	6	1.45	1.07	− 0.36	4
4	1.17	1.36	− 0.15	4	**10.00**	1.36	1.17	1	**2.00**	1.36	1.21	1
5	1.36	1.09	− 0.17	5	1.36	1.09	− 0.48	4	1.36	1.09	− 0.43	5
6	1.18	1.15	− 0.46	6	1.18	1.15	− 0.41	3	1.18	1.15	− 0.53	6
ARS				0				2.33				1.02
Panel B: min-max normalization and linear aggregation												
1	1.68	1.05	0.50	1	1.68	1.05	0.04	4	1.68	1.05	0.34	3
2	1.02	1.87	0.50	2	1.02	1.87	0.50	2	1.02	1.50	0.50	2
3	1.45	1.07	0.34	3	1.45	1.07	0.04	5	1.45	1.07	0.23	4
4	1.17	1.36	0.30	4	**10.00**	1.36	0.69	1	**2.00**	1.36	0.69	1
5	1.36	1.09	0.28	5	1.36	1.09	0.04	6	1.36	1.09	0.20	5
6	1.18	1.15	0.18	6	1.18	1.15	0.07	3	1.18	1.15	0.14	6
ARS				0				1.98				1.02

We use bold numbers to indicate the numbers that correspond to contaminated values

$$ARS = \frac{1}{n} \sum_{i=1}^{n} |Rank(CI_i) - Rank(\widetilde{CI}_i)|, \tag{1}$$

where CI_i and $\widetilde{CI}_i$ represent the composite indicator values of entity i calculated from data contaminated and non-contaminated with extreme values (or winsorized versus non-winsorized data).

2.2 Motivating Example

In what follows, we use two numeric examples to illustrate the rank shifts of composite indicators caused by extreme values in the indicators. We first present the result without winsorization and then compare it to what we obtain using the proposed multivariate distribution-driven winsorization approach.

We assume in Table 1 that we have a sample of six countries and two indicators from which to construct a composite indicator. The clean data are the reference data. In case of contamination, we generate a data error for country 4 in the first indicator. Its value of 1.17 is replaced by 10. As a result, the ranks shift compared to the clean data ranking. In panel A, the composite indicator is constructed using the so-called z-score normalization and linear aggregation methods. We explain this in more detail in Sect. 3. Note that the presence of one extreme value in the first indicator changes the composite indicator ranking to 5, 2, 6, 1, 4 and 3, instead of 1, 2, 3, 4, 5 and 6. The corresponding average rank shift is as high as 2.33. By contrast, if we use the proposed winsorization approach, the average rank shift equals 1.02. Note that the winsorization of the indicator data has an ambiguous effect on the cardinality values of the composite indicator. The winsorization leads to less extreme values in both the numerator and denominator of the normalized winsorized indicator. Hence, the aggregate effect on the cardinal value is case-specific.

In panel B, we also present a numerical example with a composite indicator constructed using the min-max normalization and linear aggregation. In this scenario, similar results to the previous situation are obtained.

3 Methodology

3.1 The Framework

Assume that we have data for n countries on p indicators. Henceforth we call the indicators sub-indicators to make a clear distinction with the composite indicator. The vector $\boldsymbol{I}^j$ consists of the n observations of the jth sub-indicator:

$$\boldsymbol{I}^j = (I_1^j, \dots, I_n^j)', \quad j = 1, \dots, p. \tag{2}$$

Throughout the paper, we use the following general definition of a composite indicator that aggregates the p standardized winsorized sub-indicators:

$$CI_i = h(g_{a,b}(w^1(I_i^1)), \dots, g_{a,b}(w^p(I_i^p))), \tag{3}$$

where $h(\cdot)$ represents the aggregation function, $g_{a,b}(\cdot)$ represents the standardization function and $w_j(\cdot)$ is the winsorization function. When $w^j(z) = z$, for $j = 1, \dots, p$, Eq. (3) nests as a special case the traditional composite indicators with no winsorization.

3.2 Choice of Aggregation and Standardization Functions

The aggregation function $h(\cdot)$ takes the form of either a weighted arithmetic average or a weighted geometric average (OECD 2008b):

$$h(z_1, \dots, z_p) = \sum_{j=1}^{p} v_j z_j \tag{4}$$

$$h(z_1, \dots, z_p) = \Pi_{j=1}^{p} z_j^{v_j}. \tag{5}$$

For both aggregation methods, the weights are required to sum to unity, *i.e.*, $\sum_{j=1}^{p} v_j = 1$. The geometric aggregation method requires in addition that all composite indicators are positive, *i.e.*, $z_j \geq 0$ in (5), for all $j = 1, \dots, p$. As standardization method $g_{a,b}(\cdot)$, z-score and min-max approaches are often used (OECD 2008b). For z-score standardization, we have

$$g_{a,b}(I^j) = \frac{I^j - a}{b}, \tag{6}$$

where $a = \overline{I^j} = \frac{1}{n} \sum_{i=1}^{n} I_i^j$ and $b = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (I_i^j - \overline{I^j})^2}$. The min-max standardization approach leads to standardized values between 0 and 1 using

$$g_{a,b}(I^j) = \frac{I^j - a}{b - a}, \tag{7}$$

where $a = \min\{I_1^j, \ldots, I_n^j\}$ and $b = \max\{I_1^j, \ldots, I_n^j\}$, n is the number of observations in each sub-indicator.

3.3 Standard Quantile-Based Winsorization

Several approaches exist to define the winsorization function $w(\cdot)$. Here, we first introduce the winsorization function based on empirical quantiles (see, e.g., Saisana 2010, for a recent application). Let $\{I_1^j, \ldots, I_n^j\}$ be a sample from $\boldsymbol{I}^j$, and let $I_{(1)}^j, \ldots, I_{(n)}^j$ be the corresponding order statistics. Then the empirical quantile of the jth sub-indicator Q^j is given by

$$\hat{Q}^j(\alpha) = I^j_{(\lceil n\alpha \rceil)}. \tag{8}$$

Based on the empirical quantile of the sub-indicator, we define the quantile-based winsorization function $w_{QF}(\cdot)$ as

$$w^j_{QF}(I_i^j) = \begin{cases} \hat{Q}^j(\alpha_1) - kIQR^j_{\alpha_1,\alpha_2} & \text{if} \quad I_i^j \leq \hat{Q}^j(\alpha_1) - kIQR^j_{\alpha_1,\alpha_2} \\ I_i^j & \text{if} \quad \hat{Q}^j(\alpha_2) - kIQR^j_{\alpha_1,\alpha_2} < I_i^p < \hat{Q}^j(\alpha_2) + kIQR^j_{\alpha_1,\alpha_2} \\ \hat{Q}^j(\alpha_2) + kIQR_{\alpha_1,\alpha_2}(I^j) & \text{if} \quad I_i^j \geq \hat{Q}^j(\alpha_2) + kIQR^j_{\alpha_1,\alpha_2} \end{cases}, \tag{9}$$

where $IQR^j_{\alpha_1,\alpha_2} = \hat{Q}^j(\alpha_2) - \hat{Q}^j(\alpha_1)$ is the inter-quantile range. k is a suitable critical value. Data outside the range $[\hat{Q}^j(\alpha_1) - kIQR^j_{\alpha_1,\alpha_2}, \hat{Q}^j(\alpha_2) + kIQR^j_{\alpha_1,\alpha_2}]$ are replaced with the border values.

We refer to this method as QF (quantile function) throughout the paper. Since the entities' values outside the range are replaced by the same border value, the entities' rankings in sub-indicators are not preserved. This limitation is addressed next by using a robust distribution-based winsorization method. We further distinguish between a univariate and multivariate approach.

3.4 Robust Univariate Distribution-Based Winsorization

The univariate distribution-driven winsorization method proceeds in two steps. First, it fits each sub-indicator with a distribution function F, based on which we identify extreme values deviating from the robust distribution. Second, it replaces the extreme values with an equidistant grid of quantiles of data from the robustly fitted distribution, thereby preserving the order of the original values.[3]

Suppose we have an estimated distribution $\hat{F}$, with quantile function $\hat{G}$. The corresponding estimators for cutoff values c_l^j, c_u^j of the jth sub-indicator are given by

[3] It should be noted that other complex distributions could be considered to model the data, but the Weibull distribution is chosen because of its simple form and because it is applicable to the indicators used in the empirical application. Inspired by Alfons et al. (2013), we replace the extreme observations with quantiles of the fitted Weibull distribution of sub-indicators.

$$\hat{c}_l^j = \hat{G}^j(\theta_1)$$
$$\hat{c}_u^j = \hat{G}^j(\theta_2). \tag{10}$$

Putting all parts together, the winsorization function $w_{DB}(\cdot)$ is as follows:

$$w_{DB}(I_i^j) = \begin{cases} \hat{G}^j(\boldsymbol{s}_L^j) & \text{if} \quad I_i^j \le \hat{G}^j(\theta_1) \\ I_i^j & \text{if} \quad \hat{G}^j(\theta_1) < I_i^j < \hat{G}^j(\theta_2) \\ \hat{G}^j(\boldsymbol{p}_L^j) & \text{if} \quad I_i^j \ge \hat{G}^j(\theta_2) \end{cases}, \tag{11}$$

where $\boldsymbol{s}_L^j = \theta_1 - (\boldsymbol{k}_L^j - 1)(\theta_1 - q_1)/(m_L^j - 1)$, $\boldsymbol{k}_L^j = [1, \ldots, m_L^j]$, $\boldsymbol{p}_L^j = \theta_2 + (\boldsymbol{k}_U^j - 1)(q_2 - \theta_2)/(m_U^j - 1)$, $\boldsymbol{k}_U^j = [1, \ldots, m_U^j]$. m_L^j and m_U^j are the numbers of observations exceeding the critical values on the left and right side, respectively, and q_1 and q_2 are the lower and upper bounds of the calibrating process. Since the winsorization function depends on the robustly fitted distribution, we refer to this method as DB (distribution-based). It uses the robustly fitted distribution for both the detection of extremes (condition in the if statement) and to set the value used in the imputation when the thresholds are exceeded.

In the simulation study and the application, all sub-indicators considered are positive-valued and their empirical distribution function is aligned with the fit from the Weibull distribution. We therefore choose F as Weibull distribution. Its cumulative distribution function is

$$F(\beta, \lambda; x) = 1 - \exp[-(x/\lambda)^{\beta}], \tag{12}$$

where β is the shape parameter and λ is the scale parameter. In this case, G is the quantile function of the Weibull distribution. As robust estimators for the parameters β and λ, we use the quantile estimators proposed by Boudt et al. (2011). They are given by

$$\hat{\beta} = \frac{\log G(\alpha_2) - \log G(\alpha_1)}{\log \hat{Q}(\alpha_2) - \log \hat{Q}(\alpha_1)}$$
$$\hat{\lambda} = \hat{Q}(\alpha)/[-\log(1-\alpha)]^{1/\hat{\beta}}, \tag{13}$$

where $\hat{Q}(\alpha)$ is the empirical α-quantile of the observations for any $0 \le \alpha \le 1$. In our simulation and application study, we set α_1, α_2 and α as 1/3, 2/3 and 1/2, respectively. For these values, Boudt et al. (2011) show that the estimator has the maximum breakdown point.

3.5 Robust Multivariate Distribution-Based Winsorization

The univariate winsorization method detects extreme observations on each sub-indicator separately and thus does not consider the correlation structure between sub-indicators. This may result in failure to identify outlying observations that are not detectable using marginals only but need a multivariate approach to be detected (see, e.g., Khan et al. 2007). To cope with these cases, we propose a multivariate winsorization approach.

The multivariate method proceeds by first filtering the extreme values in the sub-indicators by applying univariate winsorization function to the data. In this step, the correlations between sub-indicators are not considered. In the second step, the robust Mahalanobis distances of each entity based on the filtered data are calculated in order to detect the correlation outliers. Based on the minimum covariance determinant (MCD) estimator of multivariate location and scatter proposed by Rousseeuw (1984, 1985), we define the robust distance RD_i of entity I_i as

$$RD_i = \sqrt{(w_{DB}(I_i) - \hat{\mu}_{MCD})' \hat{\Sigma}_{MCD}^{-1}(w_{DB}(I_i) - \hat{\mu}_{MCD})}, \tag{14}$$

where $w_{DB}(I_i)$ is the ith observation after filtering by univariate winsorization, $\hat{\mu}_{MCD}$ is the MCD estimate of location of the sample $w_{DB}(I_1), \dots, w_{DB}(I_n)$ and $\hat{\Sigma}_{MCD}$ is the corresponding MCD covariance estimate. The MCD estimator looks for a subset of observations with smallest determinant of the sample covariance matrix. Rousseeuw and Driessen (1999) introduced a fast algorithm for computing the MCD estimator. Hubert et al. (2012) proposed an even faster algorithm that does not use random subsets but rather computes a small number of (six) deterministic initial estimators, followed by concentration steps. In this paper, we compute the MCD estimates using the deterministic algorithm as implemented in the package **rrcov** (Todorov and Filzmoser 2009) with all default parameters, in particular the size of the subsets is $h = (n + p + 1)/2$. The size of the subsets is chosen by default as $h = (n + p + 1)/2$ in order to attain maximal breakdown point. The MCD estimator is highly robust at the cost of a loss of efficiency due to using a subset of only $h = (n + p + 1)/2$ observation. In order to increase the efficiency while retaining high robustness, we use the reweighted MCD estimator in which all observations are used for which the robust Mahalanobis distance is below the square root of the 97.5% quantile of the chi-square distribution with p degrees of freedom.

It is important to note that filtered observations $w_{DB}(I)$ with the same RD value are situated on the same ellipsoid in the n-dimensional space of winsorized sub-indicators $w_DB(I_1), \dots, w_DB(I_n)$. The multivariate winsorization that we recommend to use defines a cutoff value c for the largest ellipsoid allowed and then shrinks outlying observations to a point on the boundary of that ellipsoid, as in Khan et al. (2007). We further adjust such that the univariate ranking is preserved. Altogether, this leads to the following multivariate winsorization function $w_{MD}(\cdot)$:

$$w_{MD}(I_i^j) = \begin{cases} \min\{\max[\hat{G}^j(\gamma_1), \frac{c}{RD_i} w_{DB}(I_i^j)], \hat{G}^j(\gamma_2)\} & \text{if} \quad RD_i > c \\ w_{DB}(I_i^j) & \text{if} \quad RD_i \leq c, \end{cases} \tag{15}$$

where $\hat{G}^j(\gamma_1)$ and $\hat{G}^j(\gamma_2)$ are the lower and upper bounds of the calibration, γ_1 and γ_2 are corresponding probabilities and RD_i represents the robust distance computed on $w_{DB}(I_i^j)$ and c is the high quantile of the robustly fitted Weibull distribution used for winsorizing.

The cutoff value c is ideally set to a high quantile of the distribution of the RD, which is unfortunately unknown in practice. In the simulation and application, we proxy the distribution of RD by fitting a Weibull distribution to the sample of RD using a quantile robust estimator. This then leads to setting the cutoff value c to a high quantile of the robustly fitted Weibull distribution using the quantile estimator of Boudt et al. (2011), as already discussed in Sect. 3.3.

4 Simulation Study

In this section, we compare the performance of conventional and proposed winsorization methods by means of a Monte Carlo simulation. We consider different extreme value generation mechanisms and present the result in several panels of Table 2.

4.1 Simulation Setup

We generate correlated sub-indicators by constructing a Gaussian copula with Weibull distribution as marginals. The reference marginal distributions are specified as follows: $\boldsymbol{I}^1 \sim Weibull(1, 0.5)$, $\boldsymbol{I}^2 \sim Weibull(1, 1)$, $\boldsymbol{I}^3 \sim Weibull(10, 100)$, $\boldsymbol{I}^4 \sim Weibull(100, 1000)$. We set the Spearman's rho of each pair of sub-indicators between (0.5, 1), and test to make sure that the correlation matrix is positive definite. We have the following three settings:

(1) In panel A, we contaminate sub-indicators with extreme values generated from the Uniform distributions: $O^1 \sim U(k_1 \max(\boldsymbol{I}^1), k_2 \max(\boldsymbol{I}^1)), O^2 \sim U(k_1 \max(\boldsymbol{I}^2), k_2 \max(\boldsymbol{I}^2))$, $O^3 \sim U(l_1 \min(\boldsymbol{I}^3), l_2 \min(\boldsymbol{I}^3))$ and $O^4 \sim U(l_1 \min(\boldsymbol{I}^4), l_2 \min(\boldsymbol{I}^4))$, respectively. Parameters k_1, k_2 contribute to the outlyingness of extreme values on the right tail, while the parameters l_1 and l_2 are responsible for the left tail. We set $k_1 = 2, k_2 = 3, l_1 = 0.2$ and $l_2 = 0.3$. To test stability, we also enlarge the outlyingness of extreme values by increasing k_1, k_2 to 5 and 10, respectively.

(2) In panel B, we contaminate the simulated data set with extreme values in each sub-indicator drawing from normal distribution $\mathcal{N}(\mu, \sigma)$ with $\mu = k_1 \max(\boldsymbol{I}^i)$, $\sigma = 0.1$ or $\mu = k_2 \min(\boldsymbol{I}^i)$, $\sigma = 0.001$. In this case, extreme values are generated from $O^1 \sim \mathcal{N}(k_1 \max(\boldsymbol{I}^1), \sigma_1)$, $O^2 \sim \mathcal{N}(k_1 \max(\boldsymbol{I}^2), \sigma_1)$, $O^3 \sim \mathcal{N}(k_2 \min(\boldsymbol{I}^3), \sigma_2)$ and $O^4 \sim \mathcal{N}(k_2 \min(\boldsymbol{I}^4), \sigma_2)$, respectively. We set $k_1 = 3, k_2 = 0.5$ and also test choices of $k_1 = 10, k_2 = 0.1$.

(3) In panel C, we contaminate each sub-indicator with extreme values drawing from Weibull distribution. We generate data from $O^1 \sim Weibull(k_1\beta_1, k_1\lambda_1)$, $O^2 \sim Weibull(k_1\beta_2, k_1\lambda_2)$, $O^3 \sim Weibull(k_2\beta_3, k_2\lambda_3)$ and $O^4 \sim Weibull(k_2\beta_4, k_2\lambda_4)$, respectively. $Weibull(\beta_1, \lambda_1)$,...,$Weibull(\beta_4, \lambda_4)$ represent the initial distributions of corresponding sub-indicators. In this case, we set $k_1 = 3, k_2 = 0.5$ and also run $k_1 = 10$.

The observations to be contaminated are randomly selected from the observations in the sub-indicator samples based on the contamination level ϵ. The selected observations are replaced with extreme values. We include the cases of $\epsilon = 0$, $\epsilon = 2.5\%$ and $\epsilon = 5\%$. Concerning the sample size, we compare sub-indicators from different sample sizes by generating data with $n = 200$ and $n = 2000$.

For quantile-based winsorization (QF), we set α_1, α_2 and k as 0.05, 0.95 and 3, respectively. As to distribution-based winsorization (DB), we have θ_1, θ_2 to be 0.03 and 0.97, respectively, when applying z-score standardization, while setting $\theta_1 = 0.05$ and $\theta_2 = 0.95$ when min-max standardization is implemented. In the multivariate method (MD) case, we set γ_1 and γ_2 as the corresponding quantiles on both sides of the extreme values in each sub-indicator. We set c to the 97.5% quantile of the fitted Weibull distribution.

4.2 Performance of Proposed Methods

The methods proposed are evaluated by the average rank shift described in Eq. (1). The approach takes the ranking obtained using the uncontaminated data as the correct ranking. The smaller the value of ARS, the more effective is the method.

Panel A in Table 2 presents the simulation results for the case of extreme values generated from the Uniform distribution as described above. Panels B and C in Table 2 report the simulation results with extreme values generated from normal distribution and Weibull

Table 2 Simulation results for the average rank shift under (a) no contamination, (b) level of contamination $\epsilon = 0.025$ and (c) level of contamination $\epsilon = 0.05$

	z-score & linear				Min-max and linear				Min-max and geometric			
	NT	QF	DB	MD	NT	QF	DB	MD	NT	QF	DB	MD
Panel A: $p = 4, n = 200, k_1 = 2, k_2 = 3, l_1 = 0.2, l_2 = 0.3$												
$\epsilon = 0\%$	0.0	0.0	0.2	0.8	0.0	0.0	0.4	0.8	0.0	0.0	2.2	2.2
$\epsilon = 2.5\%$	8.2	7.6	6.2	**5.2**	9.8	9.0	7.6	**5.4**	9.6	9.4	6.8	**6.6**
$\epsilon = 5\%$	14.0	13.2	11.0	**9.0**	14.6	13.8	11.8	**9.4**	17.2	17.0	11.0	**11.0**
$p = 4, n = 200, k_1 = 5, k_2 = 10, l_1 = 0.2, l_2 = 0.3$												
$\epsilon = 0\%$	0.0	0.0	0.2	0.8	0.0	0.0	0.2	0.8	0.0	0.0	2.0	2.0
$\epsilon = 2.5\%$	15.6	8.2	6.6	**5.2**	16.4	8.6	6.6	**5.2**	10.4	9.4	6.8	**6.6**
$\epsilon = 5\%$	20.8	13.8	11.4	**9.0**	18.2	13.4	11.4	**9.4**	18.6	17.2	11.0	**10.8**
$p = 4, n = 2000, k_1 = 2, k_2 = 3, l_1 = 0.2, l_2 = 0.3$												
$\epsilon = 0\%$	0.0	0.0	0.0	6.0	0.0	0.0	0.0	18.0	0.0	0.0	16.0	16.0
$\epsilon = 2.5\%$	82.0	76.0	60.0	**48.0**	114.0	100.0	84.0	**52.0**	96.0	96.0	62.0	**60.0**
$\epsilon = 5\%$	134.0	128.0	106.0	**88.0**	146.0	138.0	104.0	**104.0**	178.0	178.0	106.0	**106.0**
Panel B: $p = 4, n = 200, k = 3$												
$\epsilon = 0\%$	0.0	0.0	0.2	0.8	0.0	0.0	0.8	1.8	0.0	0.0	2.2	2.2
$\epsilon = 2.5\%$	10.8	8.6	6.8	**5.2**	11.4	9.6	8.0	**5.4**	9.0	8.8	6.6	**6.6**
$\epsilon = 5\%$	16.4	14.4	11.8	**9.2**	15.6	14.0	11.8	**9.6**	16.2	15.8	11.0	**10.8**
$p = 4, n = 200, k = 10$												
$\epsilon = 0\%$	0.0	0.0	0.0	0.8	0.0	0.0	0.8	1.8	0.0	0.0	2.2	2.2
$\epsilon = 2.5\%$	84.6	40.4	30.6	**28.8**	78.2	46.4	37.8	**27.8**	10.2	9.0	6.8	**6.8**
$\epsilon = 5\%$	121.6	71.8	56.4	**45.6**	106.2	70.2	58.4	**50.8**	17.8	15.8	11.2	**11.0**
$p = 4, n = 2000, k = 3$												
$\epsilon = 0\%$	0.0	0.0	8.0	8.0	0.0	0.0	8.0	18.0	0.0	0.0	0.0	0.0
$\epsilon = 2.5\%$	654.0	524.0	388.0	**354.0**	642.0	544.0	440.0	**370.0**	82.0	80.0	60.0	**60.0**
$\epsilon = 5\%$	1080.0	922.0	718.0	**610.0**	1012.0	900.0	738.0	**662.0**	152.0	150.0	104.0	**104.0**
Panel C: $p = 4, n = 200, k_1 = 3, k_2 = 0.5$												
$\epsilon = 0\%$	0.0	0.0	0.2	0.2	0.0	0.0	0.2	0.2	0.0	0.0	0.4	0.4
$\epsilon = 2.5\%$	17.6	10.6	8.0	**6.4**	19.0	11.0	8.4	**6.8**	7.4	5.6	5.6	**5.6**
$\epsilon = 5\%$	25.0	19.0	14.4	**11.8**	26.2	19.4	14.6	**11.8**	13.8	9.8	9.2	**9.0**
$p = 4, n = 200, k_1 = 10, k_2 = 0.5$												
$\epsilon = 0\%$	0.0	0.0	0.2	0.2	0.0	0.0	0.2	0.2	0.0	0.0	0.6	0.6
$\epsilon = 2.5\%$	18.6	12.4	8.6	**6.6**	19.2	12.8	9.0	**6.8**	8.0	5.8	5.8	**5.8**
$\epsilon = 5\%$	29.2	22.6	16.2	**13.0**	28.2	23.0	16.0	**12.6**	14.4	10.4	9.8	**9.6**
$p = 4, n = 2000, k_1 = 3, k_2 = 0.5$												
$\epsilon = 0\%$	0.0	0.0	2.0	2.0	0.0	0.0	2.0	2.0	0.0	0.0	18.0	18.0
$\epsilon = 2.5\%$	178.0	108.0	82.0	**82.0**	200.0	108.0	80.0	**80.0**	76.0	52.0	52.0	**52.0**
$\epsilon = 5\%$	272.0	194.0	150.0	**148.0**	272.0	198.0	148.0	**148.0**	138.0	98.0	88.0	**84.0**

The smallest rank shift values are shown in bold

NT not treated, *QF* quantile function-based winsorization, *DB* univariate distribution-based winsorization, *MD* multivariate distribution-based winsorization distribution, respectively. They show that the distribution-based multivariate method minimizes the average rank shifts and that the distribution-based univariate winsorization function performs better than the quantile function method. The multivariate method has the lowest rank shifts.

Of course, the presented simulation study considers only several specific contaminating distributions and levels of contamination. However, we believe that they are representative for other simulation designs, which we investigated in a larger scale simulation study.[4] In case of significantly correlated sub-indicators, we recommend the distribution-based multivariate winsorization function. This winsorization function has the best robustness for composite indicator ranking compared to the other winsorization functions considered.

5 Applications

The results of our simulation presented in Sect. 4 show that the proposed winsorization methods are effective in dampening the impact of extremes on the average rank shifts of a composite indicator. The goal of this application is to evaluate how this method affects the ranking of the UNIDO CIP and the GIP proposed by Moll de Alba and Todorov (2018, 2019). Both indicators are developed by UNIDO and use sub-indicator data that have extreme observations.

Although we cannot always know in advance that the real data are contaminated by outliers, the winsorized ranking can always reveal the extra information of the actual positions when countries lose their advantage (or show their real deficiencies) due to winsorizing the extreme values.

5.1 Ranking the Competitiveness of Countries

The CIP index was developed and introduced in 2002 by UNIDO (2002) to assess industrial performance by observing countries' ability to produce and export manufactured goods. The CIP index was revised thoroughly in 2012 (UNIDO 2013) and has since been published annually by UNIDO Statistics. For a recent analysis of the CIP index and more technical details on its computation, we refer the reader to UNIDO (2017).

The CIP index uses eight indicators to measure six dimensions of industrial performance: industrial capacity, manufacturing export capacity, economy's share in world manufacturing value added, economy's share in world manufactured exports, industrialization intensity and export quality. These eight indicators measure the international trade and domestic production of each country. Data for all indicators are available from the UNIDO statistical data portal at http://stat.unido.org (UNIDO 2019). We describe the eight indicators in Panel A of Table 3.

In the UNIDO's CIP index calculation methodology, min-max standardization is used. MHVAsh and MVAsh are averaged into a new indicator INDint (industrialization intensity); similarly, MHXsh and MXsh are averaged into MXQual (manufactured exports quality). UNIDO uses geometric aggregation to compute the final composite indicator. Putting all parts together, we have the following equation representing UNIDO's CIP method:

[4] For instance, when we extend the dimension of sub-indicators to 8 or use data generated from a Weibull distribution with other parameters, we still reach similar conclusions.

Table 3 Description of sub-indicators in CIP and GIP

Sub-indicator	Definition	Motivation
Panel A: CIP sub-indicator data		
ImWMT	Manufactured exports' share in world manufacturing trade (%)	Measures the country's impact on world manufacturing trade. The larger share in world market reflects more competitiveness
ImWMVA	Value added share in world manufacturing value added (%)	Reflects the country's impact on world manufacturing value added, which indicates the country's relative performance and impact on overall manufacturing
MHVAsh	Medium- and high-tech manufacturing value added share in total manufacturing value added (%)	Represents the technological complexity of manufacturing. The larger the MHVAsh, the more technologically complex the industrial structure of a given country and its overall industrial competitiveness are
MVAsh	Manufacturing value added share in GDP (%)	Presents the contribution of the manufacturing sector to total production
MHXsh	Medium- and high-tech manufactured exports share in total manufactured exports (%)	Indicates the technological content and complexity of exports
MXsh	Manufactured exports share in total exports (%)	Reflects the importance of manufacturing in a country's export activity
MVApc	Manufacturing value added per capita (current USD)	Represents the country's industrialization level
MXpc	Manufactured exports per capita (current USD)	Measures the manufacturing sector trade ability
Panel B: GIP sub-indicator data		
GMVApc	Green manufactured value added per capita (current USD)	Measures the country's green manufacturing capacity in producing
GMXpc	Green manufactured exports per capita (current USD)	Reflects the country's green manufacturing capacity in exports
GMVAsh	Share of green manufactured value added in total manufacturing value added (%)	Presents the country's green manufacturing impact on total manufacturing production
GMXsh	Share of green manufactured exports in total manufactured exports (%)	Indicates the share of green manufacturing impact on total manufactured exports
GEMPsh	Share of green manufacturing employment in total manufacturing employment (%)	Reflects the impact of green manufacturing on country's employment
CO2VA	CO_2 emission from manufacturing per unit of manufacturing value added (ton/USD)	Presents the impact of green manufacturing on country's environment

Table 4 Descriptive statistics of sub-indicators in CIP and GIP

Sub-indicator	Min	Mean	Median	Max	Sd	MAD
Panel A: CIP sub-indicator data						
ImWMT (%)	0.000	0.673	0.059	17.233	1.902	0.086
ImWMVA (%)	0.000	0.604	0.047	24.129	0.2521	0.065
MHVAsh (%)	0.259	24.328	21.874	78.132	17.308	20.071
MVAsh (%)	0.366	12.097	11.861	34.664	6.332	5.687
MHXsh (%)	0.000	35.183	34.047	93.798	23.136	29.129
MXsh (%)	0.167	30.820	34.133	82.974	14.977	16.318
MVApc	20.566	1953.260	678.264	20100.1	2896.66	895.952
MXpc	0.524	3204.690	613.268	31038.4	5739.75	882.520
Panel B: GIP sub-indicator data						
GMVApc	0.000	141.420	33.872	1002.660	214.166	49.537
GMVAsh (%)	0.000	5.224	4.866	15.381	3.761	4.507
GEMPsh (%)	0.000	5.606	5.438	14.362	3.801	4.578
GMXsh (%)	0.000	7.367	6.606	44.533	6.643	5.739
GMXpc	0.000	446.492	81.740	4450.150	763.981	117.822
CO2VA	0.043	0.658	0.348	3.548	0.722	0.318

$$\begin{aligned} \text{CIP} = h_1(&g(\text{ImWMT}), g(\text{ImWMVA}), h_2(g(\text{MHVAsh}), g(\text{MVAsh})), h_2(g(\text{MHXsh}),\\ &g(\text{MXsh})), g(\text{MXpc}), g(\text{MVApc})), \end{aligned} \tag{16}$$

where $h_1(\cdot)$, $h_2(\cdot)$ and $g(\cdot)$ represent geometric aggregation, linear aggregation and min-max standardization functions, respectively. The aggregation functions $h_1(\cdot)$ and $h_2(\cdot)$ take equal weights.

We analyze the impact of winsorization for the CIP data in 2016. Summary statistics are reported in Panel A of Table 4. The large gaps between the mean and median of sub-indicators ImWMT, ImWMVA, MVApc and MXpc indicate that the sub-indicators have a skewed distribution with long tails. The large differences between the standard deviation (Sd) and the median absolute deviation (MAD) also indicate that there are extreme values in these sub-indicators.

We plot the countries' ranking after multivariate winsorization against the original ranking of UNIDO CIP in Fig. 1. We label the top five countries with the largest rank changes, and these countries are China, Congo, Gabon, China Taiwan and France. Apart from these, there are two countries from the top quintile (the United States of America and Italy) and three countries from the upper-middle and middle quintiles (Luxembourg, Panama and Romania) with rank shifts larger than 3. In the lower-middle and bottom quintile, countries with rank shifts larger than 3 are Algeria, Fiji, Mongolia, Moldova, State of Palestine, Afghanistan and Iraq. These results indicate that our proposed winsorization method is practical, keeping the rankings of countries not winsorized to a range of three positions up and down.

Table 5 displays detailed information for each sub-indicator before and after winsorization for the top 10 countries according to UNIDO's CIP data for the year 2016. By construction, the ranking of the sub-indicators did not change when extreme values in each sub-indicator were calibrated. Germany still ranked first by the composite indicator.

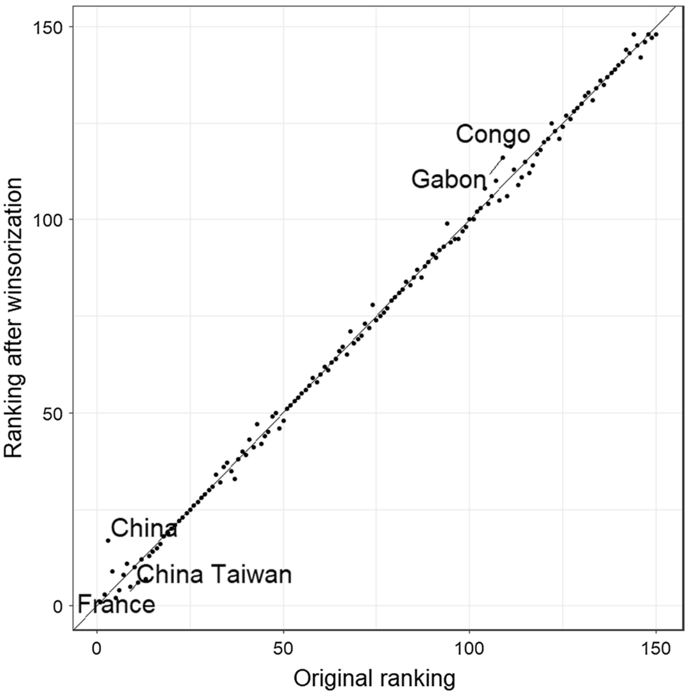

Fig. 1 Effect of winsorization on ranking of countries in UNIDO's CIP index. *Notes* Flagged countries have absolute rank shift larger than five positions. Rankings of countries below the diagonal line declined after winsorization, whereas the rankings of countries above the diagonal line rose

China's ranking changed most, dropping from third to seventeenth position. This is because China was ranked first in two sub-indicators (value added share in world manufacturing value added and manufactured exports' share in world manufacturing trade) and second on one sub-indicator (manufacturing value added share in GDP), with scores that are far ahead of other countries. After the winsorization process, these three advantages weakened. Countries with more balanced sub-indicators showed less change in ranking than China. For example, Japan's ranking changed by only one position and Belgium changed by three. The results shown in Table 5 demonstrate that our proposed winsorization method affects the ranking of countries with extreme values, while ensuring as much as possible the stability of the ranking of countries with no such extreme values. In terms of effect on the cardinal values, we see that after winsorization, all the composite indicators become larger. While winsorization has in general an ambiguous effect on the normalized sub-indicator (and hence the composite indicator), the reduction in the min-max statistic used in the normalization clearly dominates here, as compared to the effect of the winsorization on the numerator of the normalized indicator data.

The results shown in Table 5 demonstrate the importance of winsorizing extreme values in sub-indicators to weaken their impact on the ranking of treated entities. Governments and international organizations establish many policies and analyses based in part on composite indicators' ranking for extensive topics. The present example shows that an imbalance in development across indicators (i.e., a country having extremely high or low values for a given indicator) can impact the rankings in ways that are arguably disproportionate

Table 5 Winsorized CIP sub-indicator values of countries and their ranking shifts

Country	CIP	ImWMT	ImWMVA	MHVAsh	MHXsh	MVApc	MVAsh	MXpc	MXsh
Panel A: Original data result									
Germany	0.502	0.101	0.064	0.615	0.745	9659.897	0.210	14493.585	0.442
	(1)	(2)	(4)	(5)	(9)	(3)	(15)	(8)	(33)
Japan	0.384	0.049	0.092	0.562	0.813	8958.278	0.189	4566.452	0.452
	(2)	(4)	(3)	(7)	(3)	(4)	(19)	(34)	(25)
China*	0.358	0.172	0.241	0.414	0.591	2135.970	0.317	1437.617	0.480
	(3)	(1)	(1)	(27)	(28)	(41)	(2)	(57)	(7)
United States of America	0.357	0.077	0.155	0.477	0.657	6005.311	0.114	2817.100	0.370
	(4)	(3)	(2)	(17)	(17)	(15)	(77)	(42)	(67)
Republic of Korea	0.351	0.041	0.030	0.637	0.759	7331.825	0.286	9472.274	0.486
	(5)	(5)	(6)	(4)	(7)	(9)	(4)	(16)	(3)
Switzerland	0.309	0.018	0.009	0.646	0.713	13859.170	0.183	24790.820	0.342
	(6)	(16)	(17)	(3)	(12)	(2)	(22)	(4)	(75)
Ireland	0.302	0.011	0.008	0.544	0.565	20100.090	0.298	26419.807	0.473
	(7)	(25)	(22)	(8)	(32)	(1)	(3)	(2)	(12)
Belgium	0.267	0.030	0.006	0.495	0.550	6241.025	0.138	31038.463	0.443
	(8)	(9)	(28)	(15)	(40)	(13)	(51)	(1)	(32)
Italy	0.260	0.035	0.024	0.429	0.557	5018.564	0.144	6944.942	0.461
	(9)	(7)	(7)	(23)	(35)	(19)	(47)	(22)	(22)
Netherlands	0.258	0.032	0.007	0.489	0.556	5534.427	0.106	22312.253	0.426
	(10)	(8)	(23)	(16)	(36)	(17)	(85)	(5)	(46)
France*	0.256	0.037	0.023	0.501	0.670	4436.065	0.102	6715.775	0.445
	(11)	(6)	(8)	(14)	(15)	(25)	(91)	(23)	(29)
China, Taiwan*	0.244	0.023	0.009	0.689	0.750	4554.732	0.225	11479.579	0.448
	(13)	(12)	(19)	(2)	(8)	(22)	(11)	(14)	(5)
Gabon*	0.009	0.000	0.000	0.054	0.101	400.577	0.046	615.307	0.091
	(109)	(106)	(127)	(130)	(125)	(95)	(137)	(75)	(136)
Congo*	0.009	0.000	0.000	0.024	0.938	107.802	0.039	518.164	0.193
	(111)	(92)	(132)	(143)	(1)	(128)	(140)	(84)	(109)

Table 5 (continued)

Country	CIP	ImWMT	ImWMVA	MHVAsh	MHXsh	MVApc	MVAsh	MXpc	MXsh
Panel B: Winsorized data using MD method									
Germany	0.867	0.042	0.026	0.615	0.745	9659.896	0.210	13744.096	0.443
	(1)	(2)	(4)	(5)	(9)	(3)	(15)	(8)	(33)
Japan	0.703	0.041	0.028	0.562	0.813	8958.278	0.189	4566.452	0.452
	(3)	(4)	(3)	(7)	(3)	(4)	(19)	(34)	(25)
China*	0.453	0.046	0.028	0.414	0.591	2123.970	0.219	1437.617	0.481
	(17)	(1)	(1)	(27)	(28)	(41)	(2)	(57)	(7)
United States of America	0.546	0.042	0.027	0.473	0.652	5609.590	0.114	2817.101	0.370
	(9)	(3)	(2)	(17)	(17)	(15)	(77)	(42)	(67)
Republic of Korea	0.772	0.040	0.025	0.637	0.760	7331.825	0.218	9472.274	0.486
	(2)	(5)	(6)	(4)	(7)	(9)	(4)	(16)	(3)
Switzerland	0.614	0.018	0.009	0.646	0.713	9796.061	0.183	14283.360	0.342
	(4)	(16)	(17)	(3)	(12)	(2)	(22)	(4)	(75)
Ireland	0.558	0.011	0.008	0.544	0.565	10649.027	0.218	14572.429	0.473
	(8)	(25)	(22)	(8)	(32)	(1)	(3)	(2)	(12)
Belgium	0.541	0.030	0.006	0.495	0.550	6241.025	0.138	14722.314	0.443
	(11)	(9)	(28)	(15)	(40)	(13)	(51)	(1)	(32)
Italy	0.598	0.035	0.024	0.429	0.556	5018.564	0.144	6944.942	0.461
	(5)	(7)	(7)	(23)	(35)	(19)	(47)	(22)	(22)
Netherlands	0.542	0.032	0.008	0.489	0.556	5534.427	0.106	14143.882	0.426
	(10)	(8)	(23)	(16)	(36)	(17)	(85)	(5)	(46)
France*	0.581	0.037	0.023	0.501	0.670	4436.065	0.102	6715.775	0.445
	(6)	(6)	(8)	(14)	(15)	(25)	(91)	(23)	(29)
China, Taiwan*	0.565	0.023	0.009	0.689	0.750	4554.732	0.216	11479.579	0.482
	(7)	(12)	(19)	(2)	(8)	(22)	(11)	(14)	(5)
Gabon*	0.018	0.000	0.000	0.054	0.101	400.577	0.046	615.307	0.091
	(116)	(106)	(127)	(130)	(125)	(95)	(137)	(75)	(136)

Table 5 (continued)

Country	CIP	ImWMT	ImWMVA	MHVAsh	MHXsh	MVApc	MVAsh	MXpc	MXsh
Congo*	0.016 (119)	0.000 (92)	0.000 (132)	0.024 (143)	0.937 (1)	107.802 (128)	0.039 (140)	518.164 (84)	0.193 (109)

The sub-indicator ranks are given in parentheses. The top ten countries are the best performers in terms of CIP. The bottom four starred countries have the largest ranking shift due to the winsorization to the significance of these outliers, depending on the method used. Tarabusi and Guarini (2013) propose several methods for processing unbalanced indicators. Our proposed method uses winsorization functions to identify and replace the extreme values based on sub-indicators' distribution instead of using preset cutoff value. As such, it appears to offer advantages over methods presently in use, and we believe that its use should be adopted as a "best practice" in conducting analyses of this kind.

5.2 Ranking the Green Manufacturing Performance of Countries

Inspired by the UNIDO CIP index for measuring competitive manufacturing performance, Moll de Alba and Todorov (2018, 2019) construct a composite indicator that helps to gain an overall understanding of the status of green, i.e., environmentally friendly industry at the country level. The proposed indicator, called the Green Industrial Performance (GIP) indicator can be used as a complementary tool to the CIP index for analysing the inclusive and sustainable industrial development at the country level. It is comprised of six sub-indicators, namely Green manufactured value added per capita, Green manufactured exports per capita, Share of green manufactured value added in total manufacturing value added, Share of green manufactured exports in total manufactured exports, Share of green manufacturing employment in total manufacturing employment and the CO_2 emission from manufacturing per unit of manufacturing value added. The sub-indicators are quantitative, based on internationally comparable data available from recognised international data sources.

To calculate the composite indicator, each of the six indicators is normalized into the range [0,1], with higher scores representing better outcomes. Normalization is carried out by the min-max method, where the minimum and maximum values of each indicator sample values are taken. For the CO_2 emissions by manufacturing value added, for which higher values of the original indicator represents worse outcomes, we adjust the equation

$$g^*_{a,b}(I^j) = \frac{b - I^j}{b - a}, \tag{17}$$

where $a = \min\{I^j_1, \dots, I^j_n\}$ and $b = \max\{I^j_1, \dots, I^j_n\}$, and n is the number of observations in each sub-indicator. The resulting GIP index is then given by

$$\begin{aligned} \text{GIP} = h_1(&g(\text{GMVApc}), g(\text{GMXpc}), g(\text{GMVAsh}), g(\text{GMXsh}), g(\text{GEMPsh}), \\ &g^*(\text{CO2VA})), \end{aligned} \tag{18}$$

where $h_1(\cdot)$ and $g(\cdot)$ represent, respectively, geometric aggregation and the (adjusted) min-max standardization functions.

As in Moll de Alba and Todorov (2019), we compute the GIP index for the year 2015 and rank the green performance of 104 economies.

The descriptive statistics are reported in Panel B of Table 4. We find large discrepancies between the classical and robust location (mean versus median) and scale (standard deviation versus median absolute deviation) statistics. For CO2VA, we have a clear case of implosion in the normalized sub-indicator. We illustrate this in Fig. 2 where we plot the normalized original and winsorized sub-indicators versus the original data. Since the normalization uses the adjusted min-max standardization formula in (7), we have a negative dependence between the original CO2VA data and its normalized version. The extreme observations correspond to Iraq, Syria, Mongolia, Nepal, Kyrgyzstan, Kazakhstan, Viet Nam, Ukraine, China Hong Kong SAR and Oman.

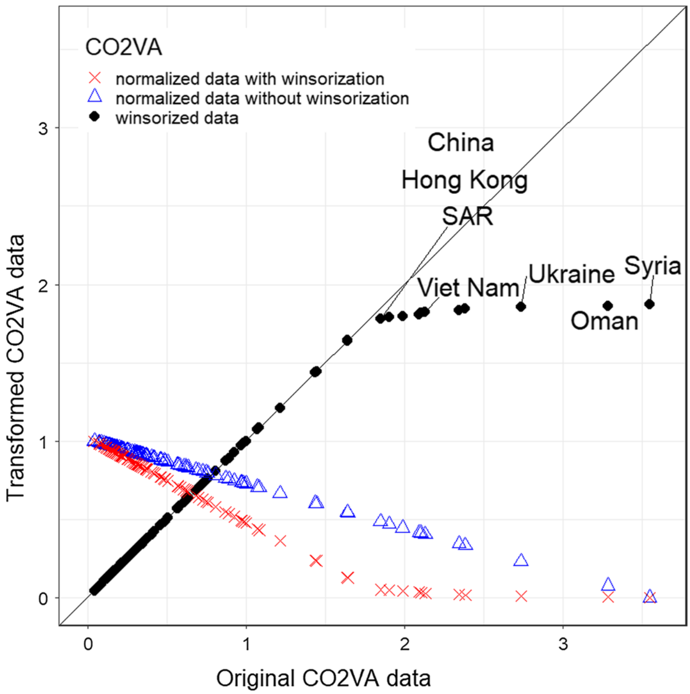

Fig. 2 Effect of winsorization and normalization on the CO2VA data. *Note*: The normalized data is negatively related to the original data since we use the adjusted min-max standardization method in (7)

We plot the countries' ranking after winsorization against the original ranking of GIP in Fig. 3. We label the three countries (or areas) with the rank shifts larger than 10 positions, namely China Hong Kong SAR, Ukraine and Viet Nam and provide more details in Table 6. Before winsorization, the bad performance on CO_2 emission of these countries was hidden and masked during the normalization by the worst one, Syria. After winsorization, they are penalized by losing positions in GIP ranking based on the true CO_2 emission performance compared to the other countries. It also happens to Oman and Kuwait (seven positions back). Trinidad and Tobago drops seven positions after the advantage on share of green manufactured exports in total manufactured exports has been winsorized.

6 Conclusions

We study the design of the winsorization method used when constructing a composite indicator for which the ranks are robust to the occurrence of extreme values in the indicators. For this purpose, we introduce two novel winsorization functions (univariate and multivariate) that are calibrated based on a robustly fitted Weibull distribution. These functions have the invariance property that they only affect the ranks of the composite indicator, while preserving the ranks at the level of the individual indicator.

The importance of winsorizing extreme values in indicators is demonstrated in an intuitive numeric experiment and an extensive simulation study. We show that, in the presence of

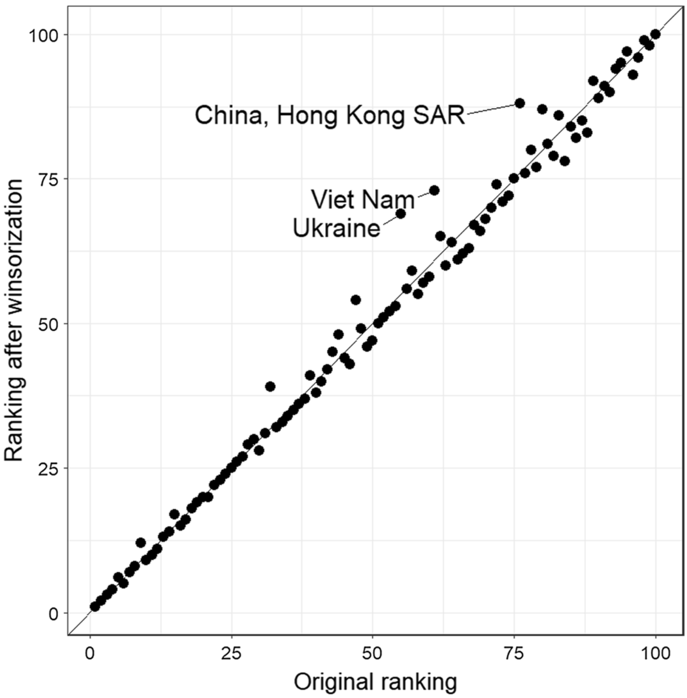

Fig. 3 Effect of winsorization on ranking of countries in GIP ranking. *Note*: Flagged countries have absolute rank shift larger than seven positions. Rankings of countries below the diagonal line declined after winsorization, whereas rankings of countries above the diagonal line rose

outlier contamination, the winsorization leads to a ranking that is substantially closer to the clean data ranking when compared to the ranking obtained using either no winsorization or the traditional winsorization based on empirical quantiles.

We then use our proposed winsorization method to solve the problem of ranking countries based on the United Nations Industrial Development Organization's Competitive Industrial Performance (CIP) and Green Industrial Performance (GIP) data. We show that the robust winsorization approach has a material impact on the ranking of the composite indicator for countries with large discrepancies in the scores of the sub-indicators. We conclude that the robustness gain of using the distribution-based winsorization method improves the reliability of composite indicator rankings for policy making. We also recommend policy makers to always conduct a sensitivity analysis of their ranking to winsorizing the data in order to evaluate potential disproportionate propagation effects of imbalanced indicator data on the composite indicator ranking.

An interesting direction for further research would be to apply the proposed distribution-based winsorization approach to other composite indicators in different application areas, such as sustainable development or human well-being.

Table 6 Winsorized GIP sub-indicator values of countries and their ranking shift

Country	GIP	GMVApc	GMVAsh	GEMPsh	GMXsh	GMXpc	CO2VA
Panel A: Original data result							
Germany	0.693 (1)	1002.662 (1)	0.138 (2)	0.135 (3)	0.137 (9)	1931.913 (7)	0.121 (12)
Denmark	0.684 (2)	675.653 (5)	0.107 (7)	0.139 (2)	0.192 (3)	2355.000 (3)	0.072 (4)
Czechia	0.611 (3)	510.959 (8)	0.136 (4)	0.132 (5)	0.143 (6)	1825.053 (10)	0.220 (34)
Singapore	0.584 (4)	643.690 (7)	0.071 (33)	0.086 (26)	0.105 (25)	4450.148 (1)	0.210 (29)
Republic of Korea	0.571 (5)	871.676 (2)	0.100 (9)	0.116 (8)	0.135 (10)	1174.813 (15)	0.216 (33)
Austria	0.568 (6)	659.454 (6)	0.100 (9)	0.095 (20)	0.129 (14)	1876.052 (8)	0.133 (13)
Hungary	0.534 (7)	382.415 (12)	0.154 (1)	0.103 (11)	0.143 (5)	1247.785 (14)	0.215 (32)
Canada	0.500 (8)	802.430 (3)	0.137 (3)	0.144 (1)	0.079 (42)	597.848 (23)	0.312 (44)
United States of America	0.476 (9)	772.129 (4)	0.102 (8)	0.117 (7)	0.131 (12)	445.714 (31)	0.211 (30)
Italy	0.476 (10)	430.731 (10)	0.109 (6)	0.098 (18)	0.128 (15)	886.187 (19)	0.119 (11)
Trinidad and Tobago*	0.262 (32)	66.140 (44)	0.024 (75)	0.014 (86)	0.445 (1)	2147.249 (4)	1.215 (88)
Oman*	0.159 (47)	110.151 (35)	0.0530 (49)	0.0729 (33)	0.111 (20)	201.142 (41)	3.284 (103)
Ukraine*	0.112 (55)	17.442 (67)	0.071 (32)	0.111 (9)	0.069 (50)	39.405 (60)	2.736 (102)
Viet Nam*	0.095 (73)	26.108 (57)	0.047 (55)	0.040 (65)	0.036 (70)	44.574 (58)	2.132 (99)
China, Hong Kong SAR*	0.058 (76)	1.914 (90)	0.003 (94)	0.003 (96)	0.057 (59)	3210.698 (2)	1.849 (94)
Kuwait*	0.049 (80)	10.579 (78)	0.005 (91)	0.011 (90)	0.031 (75)	78.307 (54)	2.106 (98)
Panel B: Winsorized data using MD method							
Germany	0.741 (1)	956.909 (1)	0.138 (2)	0.135 (3)	0.137 (9)	1931.913 (7)	0.121 (12)
Denmark	0.739 (2)	675.653 (5)	0.107 (7)	0.139 (2)	0.192 (3)	2355.000 (3)	0.072 (4)
Czechia	0.656 (3)	510.959 (8)	0.136 (4)	0.132 (5)	0.143 (6)	1825.053 (10)	0.220 (34)
Singapore	0.620 (4)	643.690 (7)	0.071 (33)	0.086 (26)	0.105 (25)	4406.797 (1)	0.210 (29)
Republic of Korea	0.599 (6)	831.852 (2)	0.100 (9)	0.114 (8)	0.133 (10)	1131.566 (15)	0.216 (33)
Austria	0.609 (5)	659.454 (6)	0.100 (9)	0.095 (20)	0.129 (14)	1876.052 (8)	0.133 (13)
Hungary	0.572 (7)	382.415 (12)	0.153 (1)	0.103 (11)	0.143 (5)	1247.785 (14)	0.215 (32)
Canada	0.523 (8)	721.972 (3)	0.137 (3)	0.144 (1)	0.079 (42)	597.848 (23)	0.316 (44)

Table 6 (continued)

Country	GIP	GMVApc	GMVAsh	GEMPsh	GMXsh	GMXpc	CO2VA
United States of America	0.499 (12)	682.711 (4)	0.102 (8)	0.117 (7)	0.131 (12)	445.714 (31)	0.211 (30)
Italy	0.510 (9)	430.731 (10)	0.109 (6)	0.098 (18)	0.128 (15)	886.187 (19)	0.119 (11)
Trinidad and Tobago*	0.211 (39)	66.031 (44)	0.024 (75)	0.014 (86)	0.289 (1)	2140.290 (4)	1.212 (88)
Oman*	0.110 (54)	110.151 (35)	0.0530 (49)	0.0729 (33)	0.111 (20)	201.142 (41)	1.864 (103)
Ukraine*	0.072 (69)	17.442 (67)	0.071 (32)	0.111 (9)	0.069 (50)	39.405 (60)	1.854 (102)
Viet Nam*	0.062 (73)	26.108 (57)	0.047 (55)	0.040 (65)	0.036 (70)	44.574 (58)	1.826 (99)
China, Hong Kong SAR*	0.024 (88)	1.914 (90)	0.003 (94)	0.008 (96)	0.057 (59)	3210.698 (2)	1.781 (94)
Kuwait*	0.026 (87)	10.579 (78)	0.005 (91)	0.011 (90)	0.031 (75)	78.307 (54)	1.817 (98)

The sub-indicator ranks are given in parentheses. The top ten countries are the best performers in terms of GIP. The bottom six starred countries have the largest ranking shift due to the winsorization

References

Alfons, A., Templ, M., & Filzmoser, P. (2013). Robust estimation of economic indicators from survey samples based on Pareto tail modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*, 271–286.

Boudt, K., Caliskan, D., & Croux, C. (2011). Robust explicit estimators of Weibull parameters. *Metrika*, *73*(2), 187–209.

Booysen, F. (2002). An overview and evaluation of composite indices of development. *Social Indicators Research*, *59*(2), 115–151.

Cherchye, L., Moesen, W., Rogge, N., & Van Puyenbroeck, T. (2007). An introduction to "benefit of the doubt" composite indicators. *Social Indicators Research*, *82*, 111–145.

Cherchye, L., Ooghe, E., & Van Puyenbroeck, T. (2008). Robust human development rankings. *The Journal of Economic Inequality*, *6*, 287–321.

Correia, C. M. (2014). Measuring creativity in the EU member states. *Investigaciones Regionales: Journal of Regional Research*, *30*, 7–26.

Davino, C., & Romano, R. (2014). Assessment of composite indicators using the ANOVA model combined with multivariate methods. *Social Indicators Research*, *119*, 627–646.

Hubert, M., Rousseeuw, P. J., & Verdonck, T. (2012). A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics*, *21*(3), 618–637.

Khan, J. A., Van Aelst, S., & Zamar, R. H. (2007). Robust linear model selection based on least angle regression. *Journal of the American Statistical Association*, *102*(480), 1289–1299.

Maggino, F. (2017). *Complexity in society: From indicators construction to their synthesis* (Vol. 70). Berlin: Springer.

Moll de Alba, J., & Todorov, V. (2018). Green industrial performance: The GIP index. *World Review of Science, Technology and Sustainable Development*, *14*(2–3), 266–293.

Moll de Alba, J., & Todorov, V. (2019). Measurement of green industrial performance: An enhanced GIP index. *International Journal of Environment and Sustainable Development* (forthcoming).

OECD. (2008). *Glossary of statistical terms*.

OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. Paris: OECD.

Parente, F. (2019). A multidimensional analysis of the EU regional inequalities. *Social Indicators Research*, *143*(3), 1017–1044.

Permanyer, I. (2011). Uncertainty and robustness in composite indices rankings. *Oxford Economic Papers*, *64*(1), 57–79.
Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*(388), 871–880.
Rousseeuw, P. J. (1985). In W. Grossmann, G. Pflug, I. Vincze and W. Wertz (Eds.), *Mathematical statistics and applications* (Vol. B, pp. 283–297). Dordrecht: Reidel.
Rousseeuw, P. J., & Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, *41*, 212–223.
Saisana, M. (2010). ELLI-index: A sound measure for lifelong learning in the EU European Commission. EUR 24529 EN.JRC, IPSC, Ispra, Italy (pp. 15–16).
Saisana, M., & Saltelli, A. (2008). Expert panel opinion and global sensitivity analysis for composite indicators. *Computational methods in transport: Verification and validation* (pp. 251–275). Dordrecht: Springer.
Saisana, M., & Saltelli, A. (2011). Rankings and ratings: Instructions for use. *Hague Journal on the Rule of Law*, *3*(2), 247–268.
Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social Indicators Research*, *81*, 65–77.
Tarabusi, E. C., & Guarini, G. (2013). An unbalance adjustment method for development indicators. *Social Indicators Research*, *112*(1), 19–45.
Todorov, V., & Filzmoser, P. (2009). An object-oriented framework for robust multivariate analysis. *Journal of Statistical Software*, *32*(3), 1–47.
UNIDO. (2002). *Industrial development report 2002/2003: Competing through innovation and learning*. Vienna: UNIDO.
UNIDO. (2013). *Competitive industrial performance index 2012/2013: The industrial competitiveness of nations: Looking back, forging ahead*. Vienna: UNIDO.
UNIDO. (2017). *Competitive industrial performance report 2016* (Vols. 1–2). Vienna: UNIDO.
UNIDO. (2019). *Competitive industrial performance report 2018*. Vienna: UNIDO.